AN OVERVIEW: THE IMPACT OF DATA MINING APPLICATIONS ON VARIOUS SECTORS

Mümine KAYA KELEŞ

Abstract: In recent years, it has become difficult to reach to the reliable information with increasing complicated, non-significant, unclear, large and raw data. The need for accurate analysis of reliable information from large data has also increased in direct proportion to the rate of data growth. The Data Mining Method, which is a statistical application, is used in any desired area to be accessed to the reliable and meaningful information. In this study, the areas where data mining methods are used were explained, a literature review about banking and finance, education, telecommunication, health, public, construction, engineering and science sectors was made, and the impact of the data mining was discussed. This study is aimed to provide a contribution to the literature eliminating the gap in the mentioned area and to bring an innovation to the applications and work in these areas.

Keywords: data mining; engineering; knowledge discovery process; science; statistic

1 INTRODUCTION

Data are raw facts or measurements that can be recorded about events and assets, and indirectly meaningful [1]. The data needs to be collected, organized, summarized, analyzed and synthesized for decision-making purposes. According to Akpınar [2], data analysis is defined as the collection, organization, modeling, and testing of information access. Data analysis requires a systematic approach involving several important steps. Data analysis is involved in many disciplines such as data mining and statistics.

Statistics are defined as the art of learning from data. In its most general form, statistics deals with the collection of data, the description of data following it, and the analysis of data that often leads to the conclusion [3]. In almost all areas of work, statistical data analysis has great importance for the results to be valid and reliable. There is no statistical method applied from the field and subject, independent of the measurement and research problem. So, it is not possible to propose a single statistical method that can provide results depending on multiple purposes in the data analysis process. For this reason, in various fields, for different purposes, there are quite a number of statistical methods defined according to different data structures.

Data mining is a scientific discipline that takes the origin of statistics. Although data mining is basically a statistical application, the methods of data mining are somewhat different from the statistical methods [4]. The most obvious difference is that, unlike data mining methods, it is not easy to analyze the large-scale data with traditional statistical methods [5]. Data mining is widely used in diverse and interdisciplinary fields. In recent years, data mining has gained a great deal of importance due to the large amount of data in different applications belonging to various fields.

The aim of this study is to discuss the applications and the trend of data mining. For this reason, in this study, it has been decided to focus on the data mining method and its applications and determine the impact of these applications on various sectors.

For this purpose, this study focuses on presenting the data mining applications with examples in the various sectors including banking and finance, education, telecommunication, health, public, construction, engineering, and science sectors. This study contains a general overview of data mining and discusses the main fields for which data mining can be applied. This study also presents the main areas of the data mining applications used: banking and finance, education, telecommunication, health, public, construction, engineering, and science sectors.

In the first part of this study, the concept of data mining is in focus. In the second part, data mining applications which have been put forward in various international disciplines are examined in general perspective. Then, in the last part of the study, comments on the effects of systematic analysis of data are presented.

2 MATERIAL AND METHODS

In this section, firstly the definitions of data mining are explained, and then the method is defined briefly.

2.1 Data mining

There are many definitions of the data mining method, which is often used by data analysts. According to Kleinberg and his colleagues [6], data mining is "an interesting pattern extraction process from the raw data". According to the Gartner Group [7], data mining is the process of "discovering new correlations, patterns and trends that are meaningful by passing a large number of data stored in a vault". According to Fayyad [8], data mining is "a systematic process that is a step in the process of information discovery and is based on algorithms to produce patterns and data analysis applications". According to another definition, data mining is the process of accessing valuable information among the mass data collected by the enterprises [9].

Data mining is a major discipline that has foundation in statistics. Data mining has emerged in order to uncover the workable data in the databases, to remove redundant data, and to achieve accurate data in the fastest possible way. Data mining makes it easier to search for a set of rules within a large amount of data in order to find some predictions about the future. It also helps extract and use valuable data from a large amount of data.

Data mining, which is a meaningful, interesting and useful process of extracting very large amount information, consists of seven steps as a part of the Knowledge Discovery Process as shown in Fig. 1. These steps are as follows: data cleaning step, in which the noisy, erroneous and inconsistent data in the data is removed and the missing data is completed; data integration step, in which a plurality of data sources from different sources are consistently combined in one source; data selection step, in which the data relating to the analysis to be performed is determined and taken from the database; data transformation step, in which the data is prepared for the data analysis methods to be used at the modeling stage and the conversion is performed; data mining step, in which intelligent methods are applied to extract data patterns; pattern evaluation step that defines the correct and interesting patterns representing the information obtained according to the measurements made; and finally, knowledge presentation step that the obtained information is presented to the user through information exploration methods.
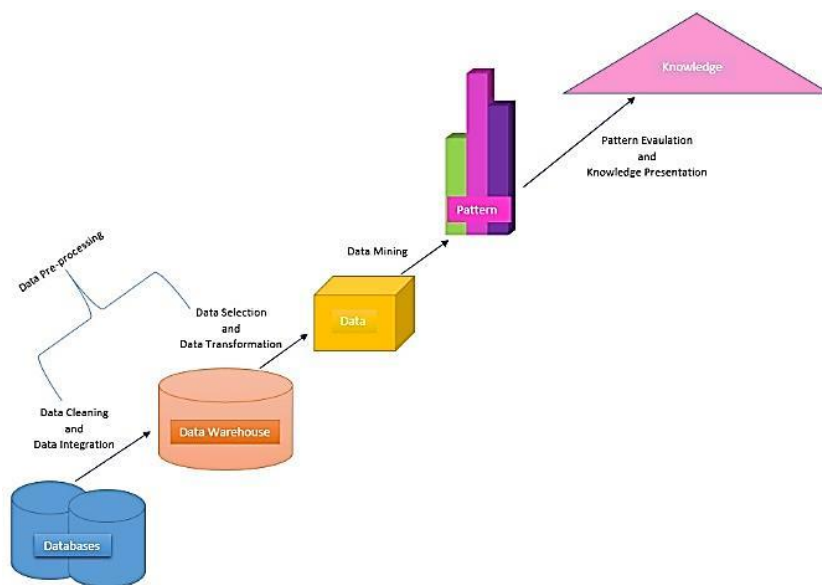

Figure 1 Knowledge Discovery (KDD) Process

2.2 Method

In this study the concepts of data and data analysis are briefly described and mentioned and the application areas of data mining are explained. The studies done in these subjects are explained and the data mining's contributions to these areas are brought into focus.

2.2.1 Data Mining Application Fields and Studies Done In These Fields

Data mining has application fields in many branches such as banking, stock exchange, marketing management, retail sales, signal processing, insurance, telecommunication, electronic commerce, health, medicine, biology, genetics, industry, construction, education, intelligence, science, and engineering [5, 10, 11]. Various businesses apply data mining to critical business processes to gain competitive advantage and help businesses grow. In this study, data mining applications in the main sectors covering the mentioned sectors were researched and explained with examples of how and for what purpose data mining was used.

2.2.1.1 Banking and Finance Sector

Data mining is mostly used in the banking and finance sectors to determine what, when, and why the customer profile prefers. At the same time, it is also used in these fields to find appropriate solutions for the right demand creation and presentation the right time demands. Furthermore, it is also used in

- financial forecasting,
- estimation of stock prices,
- management of new investments,
- determination of investment portfolio,
- formation of marketing strategies,
- making risk analyses,
- making the right choice in terms of human resources for business, credit and credit card fraud estimates,
- credit limit determination and fee management.

Data mining applications are used extensively to improve the performance of some core business processes in the banking sector [12, 13]. Some banks, for example Garanti Bank in Turkey, use data mining methods to have information about the behavior models of customers and offer appropriate and successful promotions by examining

the relationship between customer 'credit cards' selectivity and their character.

2.2.1.2 Education Sector

It has been observed that data mining has been used in many studies in the education sector including:
- determining the status of students' pass and fail,
- factors affecting the success of the students enrolling at the university,
- creating the preference of university department, determining the factors that influence the preference order of new enrolled students,
- choosing a profession according to the demographic and personal characteristics [14],
- preventing students from failing and determining the factors that affect success,
- determine the relationships between the type of school in which students graduate and their university departments [15, 16] evaluating the study activities of distance education students [16, 17],
- determining the profiles and preferences of students entering the university entrance exam [18],
- determining the relationship between academic success and participation in extracurricular activities of university students,
- determining the relationships between the socio-economic level of students and the level of academic learning [14],
- determining whether there is a relationship between student entry scores and school achievement.

These usage areas in educational sector help teachers to manage their classes, to understand their students' learning, and to provide proactive feedback to learners.

2.2.1.3 Telecommunication Sector

Data Mining can be used in the telecommunication sector:
- to predict mobile user movements in the communications sector,
- to determine the future movements of mobile users,
- to detect frauds,
- to reduce much of human-based analysis,
- to determine the factors that influence customers to call more at certain times,
- to determine user templates for social network usage [19],
- to identify new prospects using demographic data [20],
- to identify the characteristics of customers who need special action as suspension or deactivation [20],
- to prevent customer loss.

In order to prevent customer loss, telecommunication organizations can eliminate this problem by resorting to strategies for developing strategies, low cost and effective campaigns.

A good example of the use of data mining in this sector, regarding customer loss, is Verizon, America's largest wireless communications provider. In the study, Verizon

has resorted to data mining methods to identify customers whom they are likely to lose and the factors that cause customer loss.

2.2.1.4 Health Sector

Data mining techniques and application tools are more valuable for health sector. Data mining applications are used extensively to reduce the complexity of the healthcare data transactions' study in the health sector [21-24]. Data Mining is used in the health sector:
- to diagnose the disease,
- to determine the treatment method to be applied to the disease,
- to estimate the resource use and patient numbers in hospitals,
- to set the success of treatment methods applied in the hospital,
- to classify the patient data according to factors such as age, gender, race and treatment,
- to determine the high risk factors in surgeries,
- to prevent corruption in hospital expenditures.

One of the best examples of data mining studies in the health field is the one conducted at the San Francisco Heart Institute. In this study, some data such as the patient's history, laboratory data, and other medical data which are obtained from the patients to improve patient outcomes and reduce patient's hospital stay, were converted to information through data mining methods.

2.2.1.5 Public Sector

Data Mining is often used to predict public safety and security problems in the public sector. Data mining techniques offer open opportunities for the public sector to optimize decisions. These decisions are based on general trends extracted from past experience and historical data [25]. Apart from that, it is necessary:
- to determine the tax related corruption,
- to predict the impact of changes in the tax system on the budget,
- to determine waste and prevent damage caused by waste, estimate population,
- to forecast the weather, determine new job opportunities,
- to measure performance of employees, manage the business processes,
- to classify public expenditures, plan the correct use of resources,
- to forecast the future of public investment, analyze the data in defense industry,
- to determine which offenders are likely to commit crimes in terms of safety.

One of the most important examples in this area is E-Government application in Turkey. Faster feedback is being received as data mining methods are used in conjunction with the web page re-arrangement according to the behavior of the user in the past by determining the simultaneous

access of the information and the order in which the web pages on E-Government are visited.

2.2.1.6 Construction Sector

Data Mining is used in the construction sector in construction, project management, hydraulics, occupational health and safety applications, analysis of earthquake data, groundwork studies and many other areas. In view of the studies carried out in this context, it has been found that studies have been done:

- to create the information classification scheme in project documents [26],
- to determine the tax related corruption,
- to predict the impact of changes in the tax system on the budget,
- to determine waste and prevent damage caused by waste,
- to estimate the cost of highway construction [27],
- to estimate population, forecast the weather, determine new job opportunities,
- to estimate the compressive strength of the cement product [28],
- to measure performance of employees, manage the business processes,
- to define the characteristics of occupational accidents in the construction industry [29],
- to classify public expenditures, plan the correct use of resources,
- to measure worker productivity [30, 31],
- to determine the concrete compressive strength [32],
- to forecast the future of public investment, analyze the data in defense industry,
- to determine the relationship of leadership-motivation between the chief and the worker [33],
- to determine which offenders are likely to commit crimes in terms of safety,
- to determine the location of data mining method in construction management [34, 35].

One of the most important examples in this area is E-Government application in Turkey. Faster feedback is being received as data mining methods are used in conjunction with the web page re-arrangement according to the behavior of the user in the past by determining the simultaneous access of the information and the order in which the web pages on E-Government are visited.

2.2.1.7 Engineering and Science Sector

Large quantities of data have been collected from scientific fields such as astronomy, bioinformatics, computing, criminal science, engineering, geosciences, mathematics, software etc. Data Mining provides many benefits in engineering and science sector including:

- managing the process of the software used in firms,
- reducing the amount of tasks,
- increasing the speed of the Software Development Life Cycle,
- saving time and effort,

- providing competitive advantage to the organizations with the predicted analysis,
- improving the manufacturing process [20],
- biological literature analysis,
- remote sensing,
- soil quality analysis [20],
- detecting crime pattern,
- real-time feature extraction for turbulent flow analysis,
- obtaining good quality seed [20],
- evolving new crop breeds [20],
- classifying the astronomical objects,
- ecosystem modeling,
- discovering the relationships for best utilization of the cold storages and use of canal water [20],
- classifying the sequences in bioinformatics.

These usage areas in engineering and science sector help users to improve system performance of software used, to provide insight into many parts of used engineering software development processes, and to plan the future decision making process.

3 RESULTS AND CONCLUSIONS

In this study, the application areas and the contributions of data mining are examined. In the light of the literature studies examined within the scope of this study, it is observed that data mining is applied to many areas in our country. Many researchers in many different disciplines have found that meaningful results can be achieved by providing data mining methods and analysis of data. The transformation of raw data into meaningful information by processing through these studies is thought to provide economic contributions to businesses and public enterprises.

Based on these studies it can be observed that useful results can be obtained if data mining methods are used when there is an estimation or determination process in terms of cost in any sector.

It is obvious that data which are considered unnecessary and which are not recognized due to the volumetric magnitude of the preconceptions will benefit many subjects from loss of potential workforce to cost reduction, from prediction of potential customers to marketing strategies by being analyzed by data mining methods as well as statistical methods and by connecting to a system.

The important result of this study is that data mining methods, which have been used frequently in almost every field in recent years, can be applied to any kind of data obtained in the sectors mentioned in this study, and economically beneficial results can be achieved by providing competitive advantage to businesses.

4 REFERENCES

[1] Ercil Çağıltay, N., Tokdemir, G., 2010. Veritabanı Sistemleri Dersi Teoriden Pratiğe, Ada Matbaacılık, Ankara.
[2] Akpınar, H., 2014. DATA Veri Madenciliği Veri Analizi, Papatya Yayıncılık Eğitim, 1. Basım, İstanbul.

[3] Ross, S. M., 2012. Introduction to Probability for Engineers and Scientists, Academic Press.

[4] Kaya Keleş, M., 2016. İstatiksel Yöntemler ile Veri Madenciliği Yöntemlerinin Karşılaştırması (A Comparison of Statistical Methods and Data Mining Methods), Sosyal Bilimler Metinleri, 2016 Aralık ICOMEP Special Issue, pp. 20-24.

[5] Özkan, Y., 2008. Veri Madenciliği Yöntemleri, Papatya Yayıncılık Eğitim, İstanbul.

[6] Kleinberg, J., Papadimitriou, C., Raghavan, P., 1998. A microeconomic view of data mining, Data mining and knowledge discovery, 2.4, pp. 311-324.

[7] Larose, D. T., 2005. Discovering Knowledge in Data, A John Wiley & Sons, Inc., Publication, New Jersey.

[8] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, 17(3), pp. 37-54.

[9] Sümersan Köktürk, M., Dirsehan, T., 2012. Veri Madenciliği ile Pazarlama Etkileşimi, Nobel Yayın, 1st Edition.

[10] Kaya, M., Özel, S. A., 2014. Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması (Comparison of Open Source Data Mining Software), Akademik Bilişim'14 - XVI. Akademik Bilişim Konferansı (AB2014), AB2014 Proceedings, 47-53.

[11] Silahtaroğlu, G., 2008. Kavram ve Algoritmalarıyla Temel Veri Madenciliği, Papatya Yayıncılık Eğitim, İstanbul.

[12] Pulakkazhy, S., Balan R.V.S., 2013. Data Mining In Banking And Its Applications-A Review, Journal of Computer Science, 9(10), pp. 1252-1259.

[13] Bhambri, V., 2011. Application of Data Mining in Banking Sector, International Journal of Computer Science and Technology (IJCST), 2(2), pp. 199-202.

[14] Kurt, Ç., Erdem, A., 2012. Öğrenci Başarısını etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi, Politeknik Dergisi, 15(2), pp. 111-116.

[15] Ayık, Y. Z., Özdemir, A., Yavuz, U., 2007. Lise Türü ve Lise Mezuniyet Başarısının Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniği ile Analizi, Sosyal Bilimler Enstitüsü Dergisi, 10(2), pp. 441-454.

[16] Savaş, S., Topaloğlu, N., Yılmaz, M., 2012. Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 21, pp. 1-23.

[17] Çiftçi, S., 2006. Uzaktan Eğitimde Öğrencilerin Ders Çalışma Etkinliklerinin Log Verilerinin Analiz Edilerek İncelenmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü.

[18] Dolgun, M. O., Özdemir, T. G., Deliloğlu, S., 2007. Öğrenci Seçme Sınavında (ÖSS) Öğrencilerin Tercih Profillerinin Veri Madenciliği Yöntemleriyle Tespiti, Bilişim 07 Kongresi, Ankara.

[19] Bozkır, A. S., Mazman, S. G., Sezer, E. A., 2010. Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case, 2nd International Symposium on Information Management in a Changing World", Hacettepe University, Ankara, pp. 145-152.

[20] Shahbaz, M., Rahman, M., 2008. Data mining for engineering sector in pakistan: Issues and implications, Proceedings of the World Congress on Engineering and Computer Science 2008 (WCECS 2008), San Francisco, USA, ISBN: 978-988-98671-0-2.

[21] Avşar Aydın, E., Kaya Keleş, M., 2017. Breast cancer detection using K-nearest neighbors data mining method obtained from the bow-tie antenna dataset, International Journal of RF and Microwave Computer-Aided Engineering.

[22] Durairaj, M., Ranjani, V., 2013. Data Mining Applications In Healthcare Sector: A Study, International Journal Of Scientific & Technology Research, 2(10), pp. 29-35.

[23] Avşar Aydın, E., 2014. Meme Kanseri Tespitinde Mikrodalgaların Önemi ve Kanserli/Sağlıklı Meme Dokularının Yapay Zeka Algoritmaları ile Tanımlanabilmesi, Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 29(2), pp. 27-38.

[24] Aydın, A., AVŞAR AYDIN, E., 2017. Evaluation of Limestone Layer's Effect for UWB Microwave Imaging of Breast Models Using Neural Network, Technical Journal, 11(1-2).

[25] Wang, J, Hu, X., Zhu, D., 2008. Data Mining in Public Administration, Handbook of Research on Public Information Technology, IGI Global, Chapter LI, pp.556-567.

[26] Caldas, C., Soibelman, L., Han, J., 2002. Automated Classification of Construction Project Documents, J. Comput. Civ. Eng., 16(4), pp. 234-243.

[27] Wilmot, C. G., Cheng, G., 2003. Estimating Future Highway Construction Costs, J. Constr. Eng. Manage, 129 (3), pp. 272-279.

[28] Baykasoğlu, A., 2005. Veri Madenciliği ve Çimento Sektöründe Bir Uygulama, Akademik Bilişim Konferansları, pp.82-83.

[29] Liaoa, C. W., Perng, Y. H., 2008. Data Mining for Occupational Injuries in The Taiwan Construction Industry, Safety Science, 46(7), pp. 1091-1012.

[30] Kaya, M., Keleş, A. E., Laptalı Oral, E., 2013. Construction Crew Productivity Prediction by Using Data Mining Methods, Proceedings of the 4th World Conference on Learning, Teaching and Educational Leadership, Procedia - Social and Behavioral Sciences, 141, pp. 1249–1253.

[31] Keleş, A. E., Kaya, M., 2014. Duvar İnşa Edilmesinde Verimliliği Etkileyen Faktörlerin Apriori Veri Madenciliği Yöntemi Kullanılarak Analizi (The Analysis of the Factors Affecting the Productivity in the Wall Construction of the Using Apriori Data Mining Method), Akademik Bilişim Konferansları, AB 2014 Proceedings, pp.831-836.

[32] Özel, C., Topsakal, A., 2014. Veri Madenciliği Kullanarak Beton Basınç Dayanımının Belirlenmesi, Cumhuriyet Üniversitesi Fen Fakültesi Fen Bilimleri Dergisi, 35(1), pp. 43-57.

[33] Keleş, A. E., 2016. İnşaat Projelerinde Şantiye Şeflerinin Liderliği ve Çalışan Motivasyonu İlişkisinin Veri Madenciliği ile Belirlenmesi. Çukurova Üniversitesi Fen Bilimleri Enstitüsü PhD Thesis, Adana.

[34] Keleş, A. E., 2016. İnşaat Sektöründe Veri Madenciliği Uygulamalarına Genel Bakış Ve Ekonomik Etkilerinin Yorumlanması, Balkan Journal of Social Sciences, International Congress of Management Economy And Policy Special Issue, pp. 55-61.

[35] Kaya Keleş, M., Keleş, A. E., 2017. Veri Madenciliği Uygulamalarının ve Sezgisel Optimizasyon Algoritmalarının Yapım Yönetimindeki Yeri (The Place of Data Mining Applications and Heuristic Optimization Algorithms in Construction Management), Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 32(1), pp. 235-242.

Author's contact:

Mümine KAYA KELEŞ, Assistant Prof. Dr.
Adana Science and Technology University,
Department of Computer Engineering
Seyhan/Adana-TURKEY, +90-0538 548 35 00
mkaya@adanabtu.edu.tr, muminekayakeles@gmail.com